

Warum sollten sich Bibliotheken mit dem Semantic Web beschäftigen?



europæana

Kultur. Denken

Prof. Dr. Stefan Gradmann

Humboldt-Universität zu Berlin / School of Library and Information Science

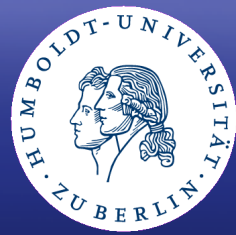
stefan.gradmann@ibi.hu-berlin.de

Übersicht



- **Warum dieser Vortrag mit dieser Titelfrage?**
- **Befassung mit dem 'Semantic Web' – Schlechte Gründe ...**
- **... und gute Gründe ...**
- **Eine 'Drohung', eine Perspektive und ein Epilog**

Warum dieser Vortrag / Warum dieser Titel?



- Wen muss ich hier noch überzeugen? Ist dieser Vortrag nicht **völlig obsolet**?
- Leider nein!
 - Ein paar Worte zur Genese ...
 - Library 101 - 101 Resources & Things to Know (RTK) ist voller Social-Software-Angebote – aber das Attribut 'semantic' **taucht dort nicht 1 Mal auf!**
 - Wohl in einer Leserreaktion: „To the list of future skills, I'd add understanding the basics of the semantic web and what linked data is. *Not that you need to know how to build triple stores and encode in RDFa, but just know the basic about the technology and the affordances it may offer information online.*“ (<http://www.libraryman.com/blog/101rtk/#section3>)
 - Das war so vor einem Jahr – **und bis heute hat sich nichts geändert!**

**... und außerdem gibt es leider auch
schlechte Gründe,
sich mit dem Semantic Web zu
beschäftigen**

Schlechte Gründe (1)

- *... weil man das heutzutage in Förderanträge schreiben muss.*
 - Der schlechteste Grund (weil wirklich gefährlich!): dies ist die beste Methode, 'Semantic Web' als Begriff und die gesamte Thematik **schnell und nachhaltig zu verbrennen!**
 - Zudem ist hierbei die Gefahr groß, **peinlichen Unfug zu schreiben**
 - Von Client-Server ...
 - ... zu „Kleinserver“!
 - Und es geht noch schlimmer ...: vom Data-W...house

Schlechte Gründe (2)

- ... weil nach Web 2.0 Web 3.0 kommt
 - Ist so problematisch wie im Grunde schon die Verwendung von 'Web 2.0' selbst.
 - 2.0 impliziert eine nicht-funktionierende Version 1.0.
 - Die Reihung 1.0, 2.0, 3.0 suggeriert eine **lineare Kontinuität**, die so nicht existiert.
 - 2.0 ist **kein Begriff**, sondern weit eher eine **Marketingvokabel**.
 - Die unausweichliche Diskreditierung der Marketingvokabel kontaminiert alles, was mit ihr in Verbindung gebracht wurde!
 - → auch '3.0' ist in diesem Sinne **gefährlich**!

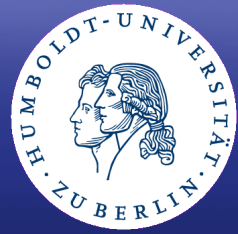
**... es gibt aber zum Glück mehr
gute Gründe,
sich mit dem Semantic Web zu
beschäftigen**

Gute und sehr gute Gründe, sich mit dem Semantic Web zu befassen



1. ... weil es sonst andere tun
2. ... um nicht im unguten Sinne museal zu werden
3. ... weil Bibliotheken nach Einführung von RDA gar nicht mehr anders können!
4. ... weil Bibliotheken auf Basis des Semantic Web ihren Benutzern attraktive und neuartige Funktionen anbieten können
5. ... weil sie einen spezifischen Beitrag zum Semantic Web leisten können
6. ... damit sie auch in Zukunft überhaupt noch wahr- und ernstgenommen werden, und zwar als Teile von 'Wissensarchitekturen'

Gute Gründe (1): Institutioneller Bestandsschutz



- ... *weil es sonst andere tun*
 - Informatiker
 - Theseus!
 - Information Professionals aus Nicht-Kultureinrichtungen
 - DGI!
- Der Grund ist ein **guter**, weil institutioneller Bestandsschutz eine starke Motivation abgibt.
- Der Grund ist dann ein **schlechter**, wenn nicht weitere Motivationen hinzukommen.

Gute Gründe (2)

- ... *um nicht im unguten Sinne museal zu werden*
 - Es besteht die reale Gefahr, dass Bibliotheken **gar nicht mehr** oder **nur als Ansammlung musealer Ressourcen** wahrgenommen werden.
 - Dem kann nur begegnet werden, wenn mit den 'Beständen' von Bibliotheken **gearbeitet** wird.
 - Zumindest wissenschaftliches Arbeiten wird dabei zunehmend auf **digitalen, RDF-basierten Heuristiken** und damit (auch) Techniken des Semantic Web beruhen.
 - → Wir tun gut daran, unsere Metadaten für diesen neuen Arbeitskontext vorzubereiten: **der erforderliche Vorlauf ist lang und aufwändig!**

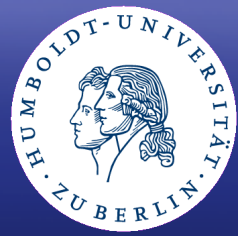
Gute Gründe (3a)

- ... weil Bibliotheken nach der anstehenden Einführung von RDA wohl kaum mehr anders können!
- Hintergrundmaterial
 - Gordon Dunsire: UNIMARC, RDA and the Semantic Web (IFLA 2009):
www.ifla.org/files/hq/papers/ifla75/135-dunsire-en.pdf bzw.
auch LIDA 2009:
<http://cdlr.strath.ac.uk/pubs/dunsireg/LIDA2009.pps>
 - Alexander Haffner: RDA and the Semantic Web (DC 2009):
www.slideshare.net/ah641054/rda-and-the-semantic-web
 - The FRBR Blog: <http://www.frbr.org/categories/sw>
 - NGC4LIB on Tim Berners-Lee and the Semantic Web:
<http://tiny.cc/NGC4LIBTBL>

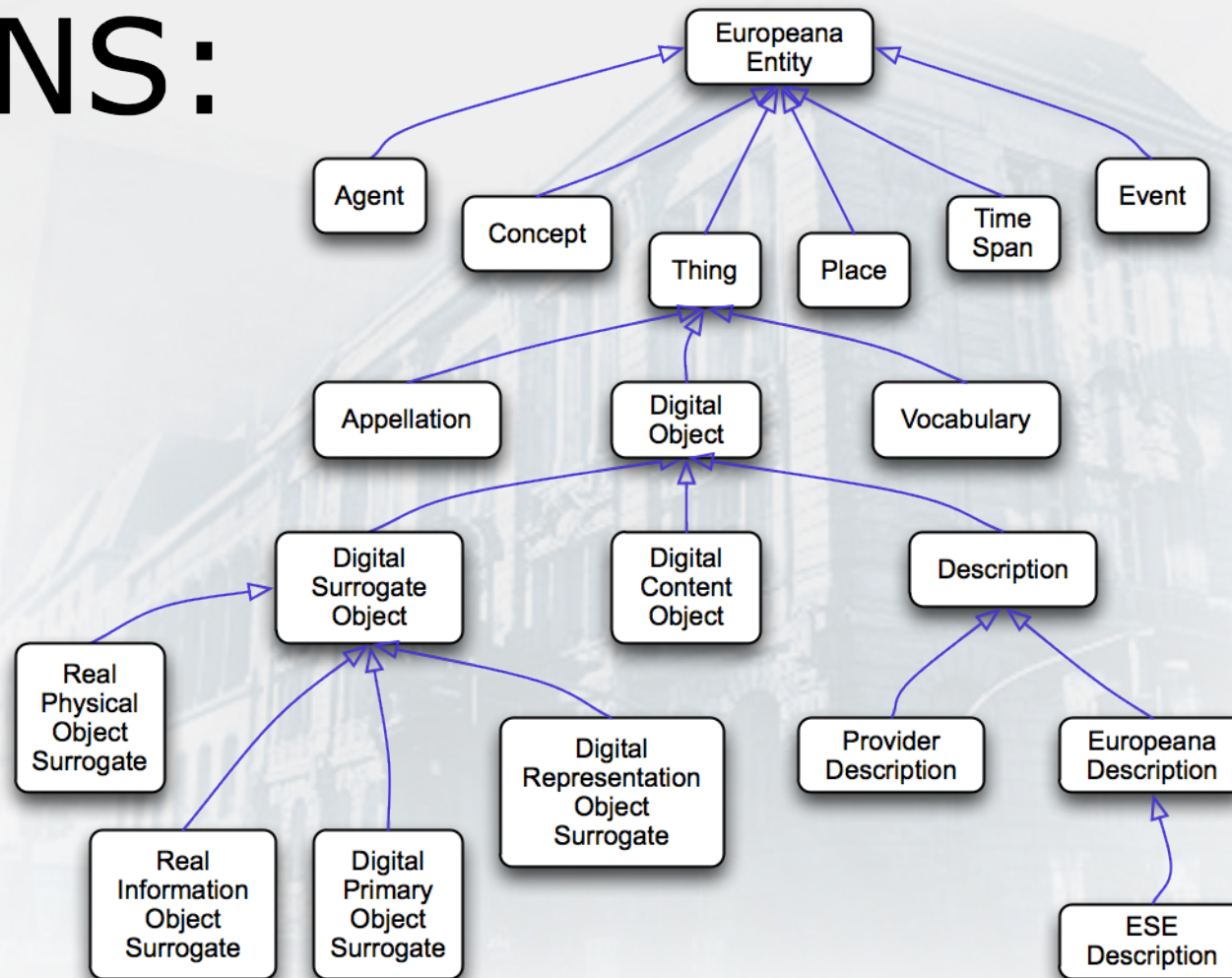
Gute Gründe (3b)

- ... weil zumindest Bibliotheken nach der anstehenden Einführung von RDA wohl kaum mehr anders können!
- Entscheidend sind die Arbeiten der **DCMI / RDA task group**, die in zwei Vokabularen mündet
 - RDA Metadaten-**Entitäten** (Elemente und Attribute) **in RDFS**
 - RDA **Vokabular** in **SKOS**
 - → <http://metadataregistry.org/rdabrowse.htm>
- RDA ist für Bibliothekare die Chance, **Teil des Informationsraumes WWW** zu werden
 - Nicht mehr nur WWW-Schnittstellen für OPAC-Monolithen
 - Wir werden Teil des WWW 'Mainstream'
 - Keine proprietären Standards mehr!
 - Bibliothekarische Standards als Spezialisierung von W3C-Standards
- NSDL Registry ↔ BIBO (<http://bibliontology.com/>)?

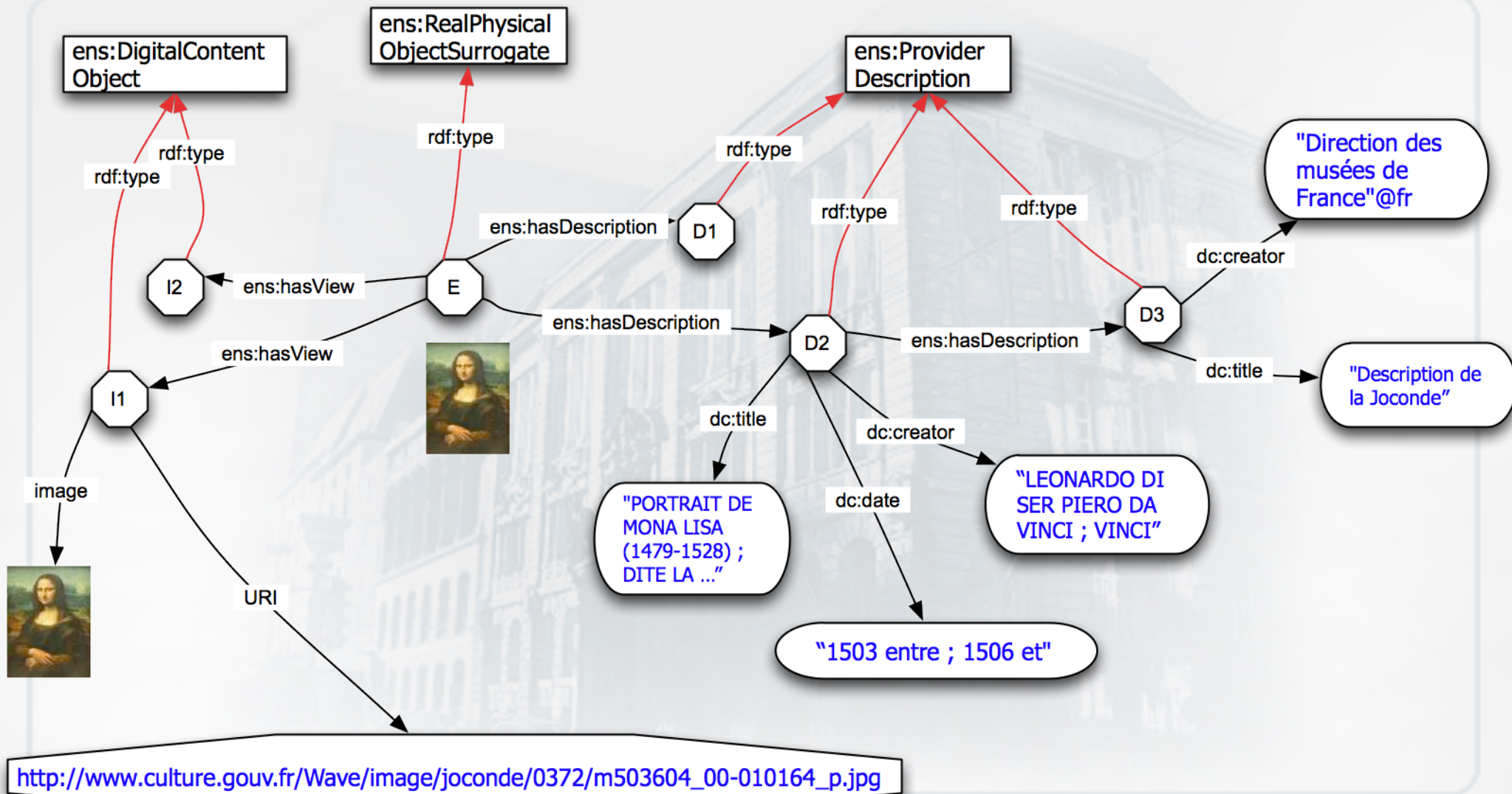
Exkurs: Keine verzichtbaren eigenen Standards! (1)



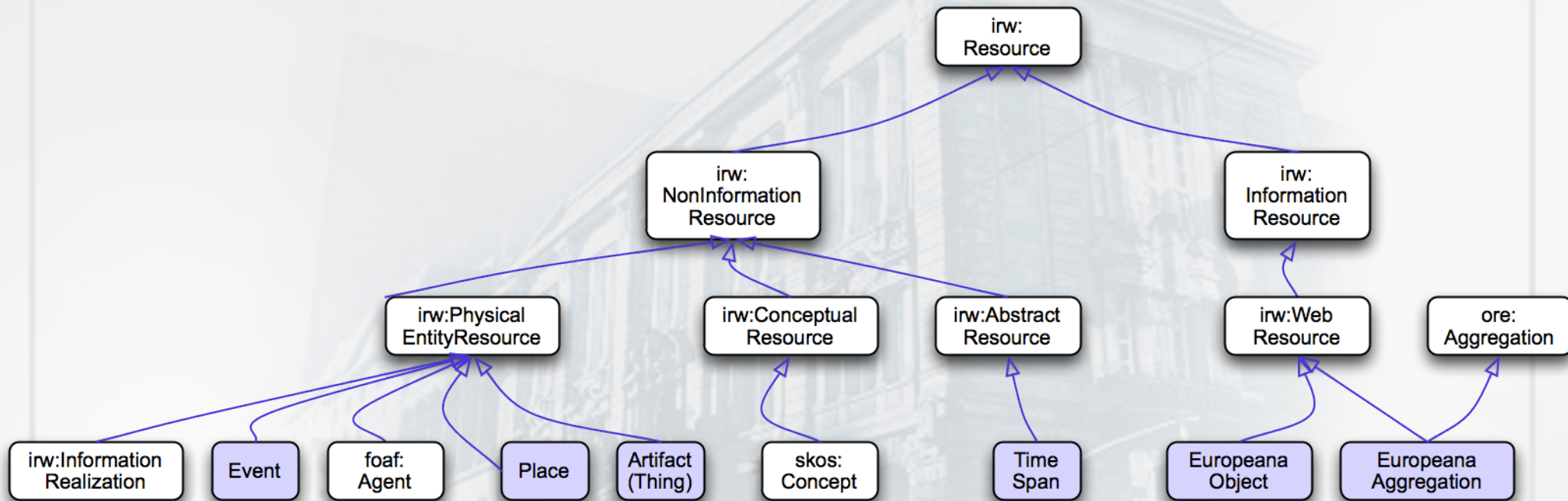
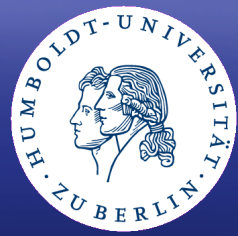
ENS:



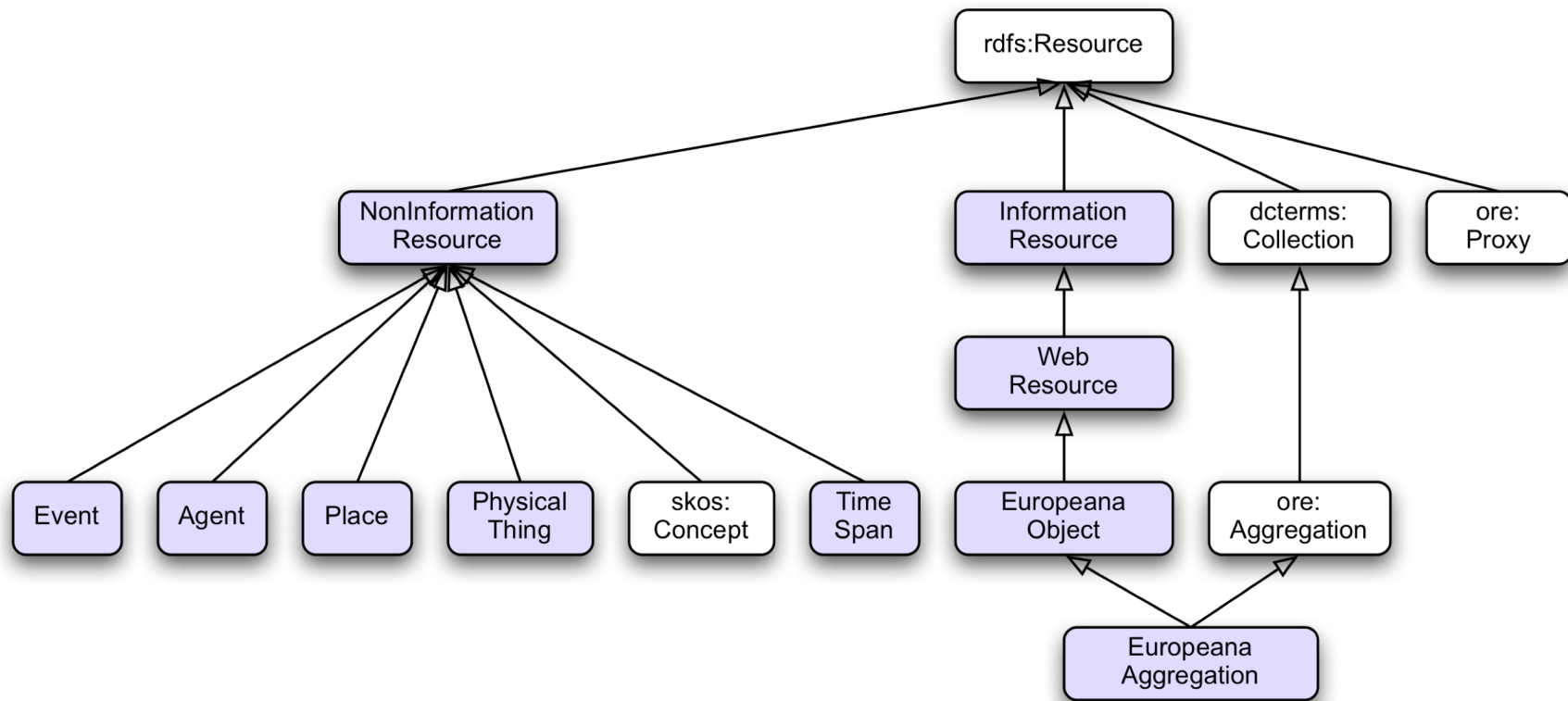
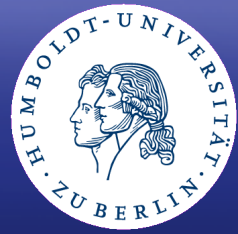
Exkurs: Keine verzichtbaren eigenen Standards! (2)



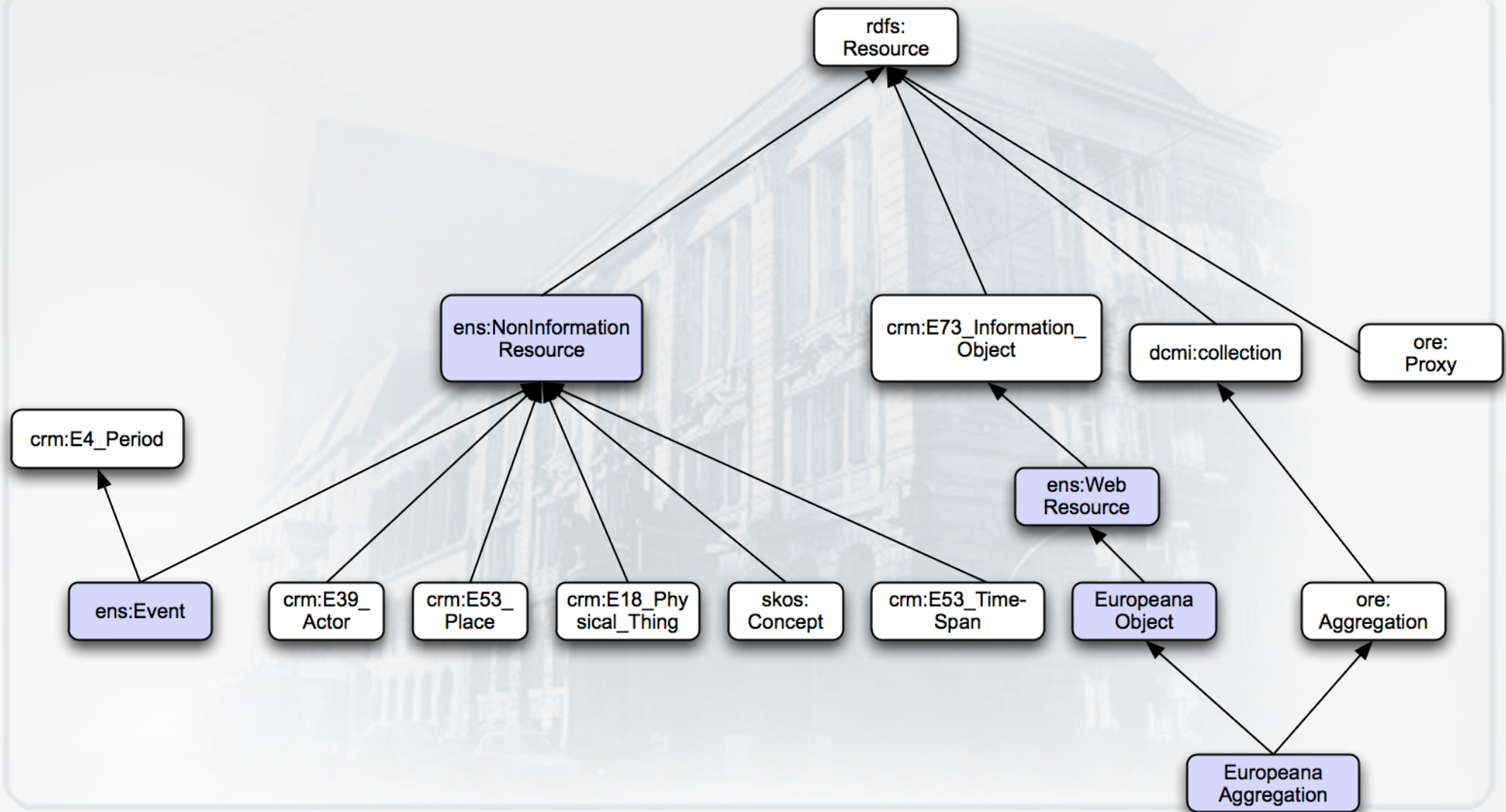
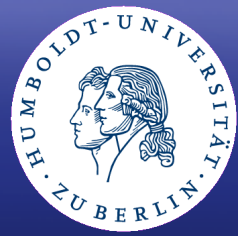
Exkurs: Keine verzichtbaren eigenen Standards! (3)



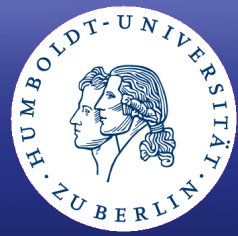
Exkurs: Keine verzichtbaren eigenen Standards! (4)



Exkurs: Keine verzichtbaren eigenen Standards! (5)



Exkurs: Von 'Records' zu 'Graphen'



Digitale
Repräsentationen

ens:WebResource
http://www.culture.gouv.fr/image/joconde/0372/m503604_00-010164_p.jpg

ens:WebResource
http://www.culture.gouv.fr/image/joconde/0372/m503604_00-010164_p.jpg

ore:aggregates
ens:hasView

ore:aggregates
ens:hasView

ore:Aggregation
[ex1:aggregation/000PE025604](#)

ore:proxyIn

"LEONARDO DI SER PIERO
DA VINCI, VINCI, Léonard de (dit)"@fr

dc:creator

ore:Proxy
[ex1:proxy/000PE025604](#)

ore:aggregates
ens:aggregatedCHO

ore:proxyFor

dcterms:title

Aggregation

ex1:schema/painting
ex1:object/000PE025604

Reales Objekt

Proxy

"PORTRAIT DE MONA LISA
(1497-1528) ; DITE LA
JOCONDE"@fr

Gute Gründe (4a)

- ... weil sie auf Basis des Semantic Web ihren Benutzern attraktive und neuartige Funktionen anbieten können
- **Inferenzbasierte Suche** (kann mehr als Volltextsuche!)
 - Gegeben seien
 - Berlin Alexanderplatz : die Geschichte vom Franz Biberkopf / Döblin, Alfred. - Frankfurt am Main : S. Fischer, [2004]
 - A moveable feast / Hemingway, Ernest. - London : Vintage, 2000 (mit SW Paris)
 - Ulysses / Joyce, James. - London : The Bodley Head, 2008 (mit SW Dublin)
 - Volltextsuche nach „Europäische Hauptstädte in der Literatur des 20. Jahrhunderts“ findet keinen dieser Datensätze.
 - Die Kombination von „Berlin CapitalOf Germany“, „Paris CapitalOf France“ und „Dublin CapitalOf Ireland“ sowie „Germany PartOf Europe“, „France PartOf Europe“ und „Ireland PartOf Europe“ würde weiterhelfen!

Gute Gründe (4b)

- ... weil sie auf Basis des Semantic Web ihren Benutzern attraktive und neuartige Funktionen anbieten können
- **Suchtermoptimierung** / Semantische Erweiterung von Suchräumen
 - Welche der drei folgenden Erweiterungen ist Unsinn – und warum?
 - Sankt Petersburg Museum → Leningrad Museum – ↗
 - iPod Hörschaden → Walkman Hörschaden – ↗
 - Leningrad Cowboys → Saint Petersburg Cowboys – ↘
 - „iPod IsA Music Device“, „Walkman IsA Music Device“ und „Leningrad RenamedTo Sankt Petersburg“
 - „Leningrad Cowboys“ ist eine 'Named Entity'
 - Regel: Termexpansion in 'Named Entities' ist nicht erlaubt.

Gute Gründe (5a)

- ... weil sie einen spezifischen Beitrag zum Semantic Web leisten können (Tim Berners Lee im Jahr 2000)



Killer App for the SWeb

Too general to be aimed at any specific application

- Ontologies - early adopters
- Catalogs on the web
- Electronic commerce transactions
- Operating across many apps (eg PDAs)
- Your favorite

A killer ap is what led you to discover a technology?

Gute Gründe (5a)

- *... weil sie einen spezifischen Beitrag zum Semantic Web leisten können*
 - RDF als Kerntechnologie des SemWeb **hat bibliothekarische Wurzeln**
 - Brays erste Einführung in RDF (1998) spricht noch von „divine metadata for the web“!
 - **Wir haben Metadaten** – die meisten WWW-Ressources haben keine (oder wenige und unzuverlässige)
 - Wir haben **Kontextualisierungsdaten**!

Gute Gründe (5b)

- ... weil sie einen spezifischen Beitrag zum Semantic Web leisten können
- Kontextualisierungsdaten:
 - Termlisten
 - Normdaten wie PND, GKD, VIAF ...
 - Klassifikationen / Schlagwörter
 - LCSH
 - DDC
 - SWD
 - Relationierungsressourcen
 - Thesauri
 - Ontologien wie MarcOnt, FRBR oder Bibliographic Ontology
- Für viele dieser Ressourcen existiert ein Migrationspfad: SKOS

Gute Gründe (5c)

- *... weil sie einen spezifischen Beitrag zum Semantic Web leisten können*
 - Das W3C setzt auf Input insbesondere von Bibliotheken – und hat ihnen mit SKOS sogar einen eigenen Standard gewidmet, denn Ziel von SKOS ist gemäß der „Reference“:
 - “... making the **accumulated experience and wisdom of knowledge organization in the library and information sciences**
 - accessible,
 - applicable within and
 - transferable to the technological context of the Semantic Web,
 - in a way that is complementary to existing Semantic Web technology (and in particular formal systems of knowledge representation such as OWL)”
 - Wann hat uns jemand zuletzt solche Komplimente gemacht?

Gute Gründe (5d)

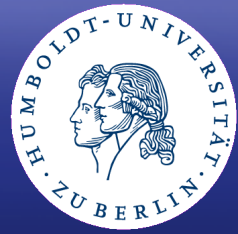
- ... weil sie einen spezifischen Beitrag zum Semantic Web leisten können – **und weil das so einfach ist!**
 - Bibliothekare müssen es nur **wollen!**
 - Die Ressourcen **existieren** in Bibliotheken und werden im Linked Data Web dringend **gebraucht!**
 - Die erforderliche Technologie existiert und **ist kein Hexenwerk**: viel mehr als RDF(S) und SKOS ist erst einmal gar nicht erforderlich
 - Erforderlich ist allerdings ein Umdenken, das Züge eines Kulturbruchs hat! **Wir sind nicht mehr der Nabel der Informationswelt!**
 - Erforderlich ist in der Folge auch eine Reform der **Ausbildungsinhalte** und des **Gremien(un)wesens**

Vielleicht der beste Grund (6a)



- ... damit sie auch in Zukunft überhaupt noch wahr- und ernstgenommen werden, und zwar als Teile von 'Wissensarchitekturen'
 - Bibliotheken waren über Jahrhunderte **akademische Leitinstitutionen**
 - Sie sind in den letzten Jahrzehnten auf eine Rolle als **Informationsdienstleister** reduziert worden
 - Diese Rolle ist in letzter Konsequenz **durchaus riskant** ...
- Was geschieht beim Übergang aus einer 'Informationsgesellschaft' in eine 'Wissensgesellschaft'?
 - Wissen = Information + Kontext
 - Bibliotheken verfügen über **Kontextualisierungsressourcen**

Vielleicht der beste Grund (6b)



- ... damit sie auch in Zukunft überhaupt noch wahr- und ernstgenommen werden, und zwar als Teile von 'Wissensarchitekturen'
 - Wenn Bibliotheken neben den Kontextualisierungsressourcen selbst auch die **Methoden** anbieten, Informationen mit diesen zu verknüpfen, haben sie eine echte Chance!
 - Diese Methoden sind diejenigen des '**Social Semantic Web**'
 - Siehe dazu die ausführliche Einführung von Sebastian Kruk <http://tiny.cc/TutorialKruk>
 - Bibliotheken werden deshalb **nicht zu Entwicklern** ...
 - ... wohl aber zu **hochkompetenten Anwendern** dieser Methoden!

Eine 'Drohung' ...

- "The Semantic Web may eventually lead to the extinction of stored catalogue records
 - Replaced by a completely distributed "one record fits all"
 - Equivalent to a single, but complex, RDF graph linking all metadata on the Web." (Gordon Dunsire)
- Einige identifikationsstiftende Kernbegriffe werden wir auf dem Weg in das 'Semantic Web' wohl aufgeben müssen
 - Katalog
 - Bestand
 - Sammlung
 - ...
- **Aber ist das wirklich so schlimm?**

... und eine Perspektive

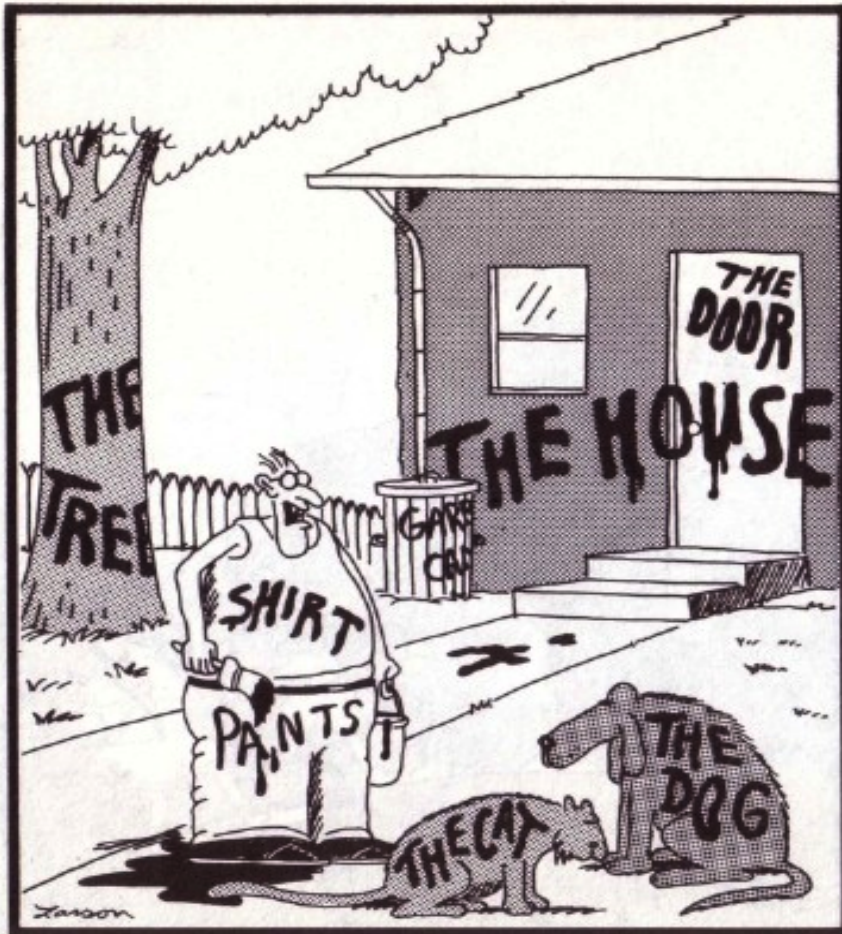
Es gibt einen zusätzlichen, strategisch bedeutsamen Beitrag von 'Gedächtnisinstitutionen' zum Semantic Web

- Das Semantic Web ist heute weitgehend **statisch** und kann mit **Bedeutungswandel** nicht adäquat umgehen
- 'Gedächtnisinstitutionen' wissen viel über Bedeutungswandel und die **Bedingungen der Langzeitverfügbarkeit von Informationen und Wissen**.
- Wir sollten dies Wissen in die 'Community' des Semantic Web einbringen.
- Dies wäre ein für alle Seiten profitabler Schritt und eine echte Perspektive für Bibliotheken!

Wie nennen wir das Kind?

Epilog zur Begrifflichkeit

Web of Things?

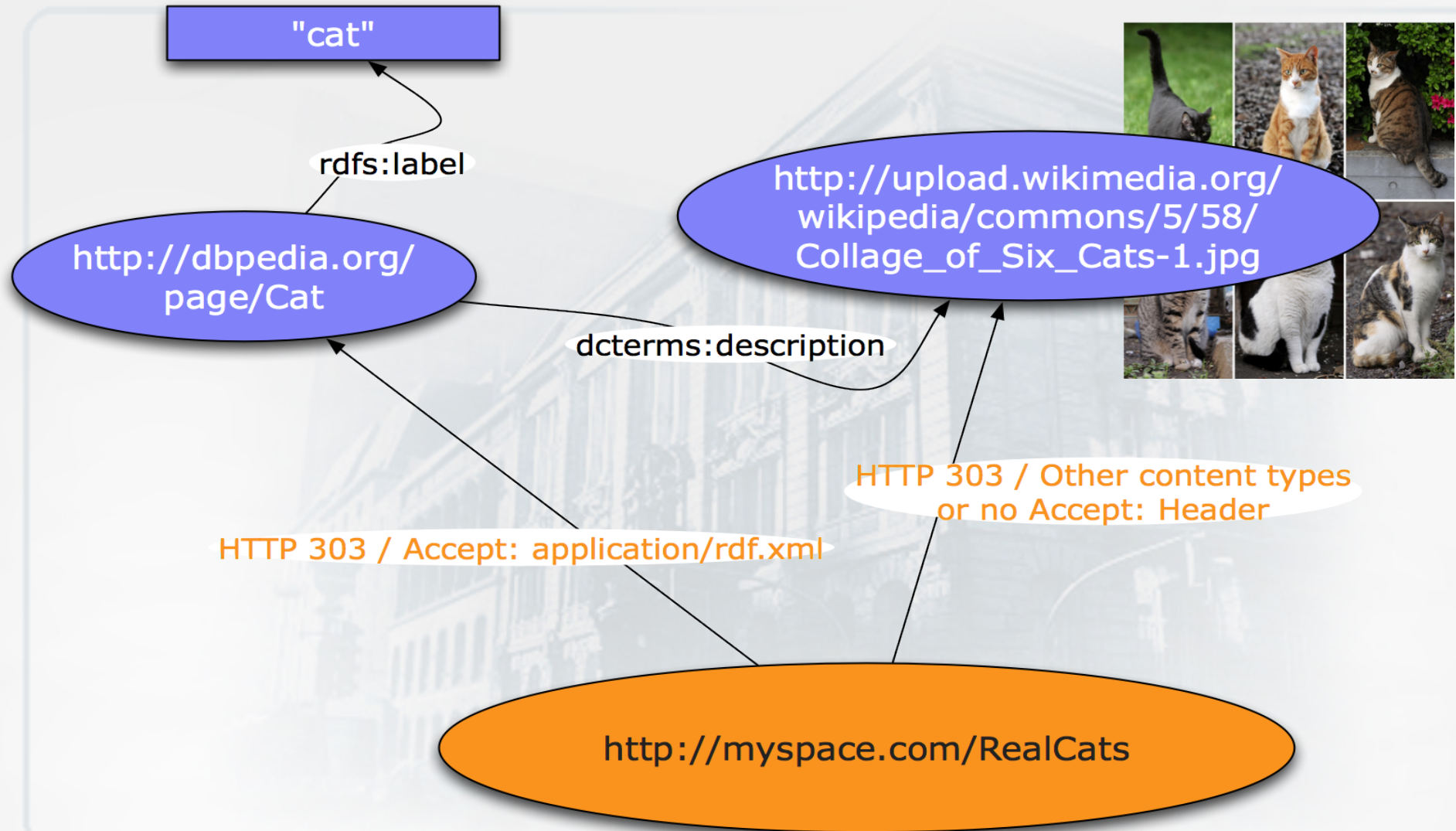


"Now! ... That should clear up
a few things around here!"

Was stimmt an
diesem Bild
nicht?

Aus dem Blog von Ronald Carpentier
[http://carpentier.wordpress.com/
2007/08/08/1-2-3/](http://carpentier.wordpress.com/2007/08/08/1-2-3/)

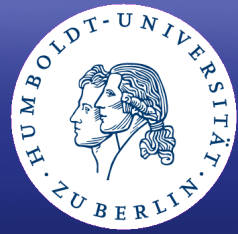
... wohl doch eher so:



Linked Open Data

- LoD: "Semantic Web **done right**" (TBL, [http://www.w3.org/2008/Talks/0617-lod-tbl/#\(3\)](http://www.w3.org/2008/Talks/0617-lod-tbl/#(3)))
- → Was war denn nun bis 2007 **falsch** am Semantic Web?
 - Das Erbe der KI (agentenbasierte Ansätze mit komplexer Logik überwogen)
 - Zumeist abgeschlossene Firmenanwendungen
 - Kaum Präsenz im WWW ("Where's the Web in the SW?" Frank van Harmelen, 2006)
 - Mißbrauch des Attributs 'semantisch'
- „I called this graph the Semantic Web, but maybe it should have been **Giant Global Graph!**" (TBL, <http://dig.csail.mit.edu/breadcrumbs/node/215>)

SemWeb, 3.0, SW, LoD → ?



- Sollten wir also lieber über **Linked Open Data** sprechen?
- Sympathischer – aber auch nicht wirklich gut ...
 - Data?
 - Open??
- 'Linked Information Web' wäre vielleicht besser ...
- Oder gleich: **Docuverse** (Ted Nelson: Literary Machines)
- → Wie auch immer wir das Kind nennen: es wird uns in den kommenden Jahren heftig beschäftigen!
- Und der Name ist wichtig: geht es doch um **die Zukunft der Bibliotheken!**

Danke für Ihre Aufmerksamkeit